Marked point process model for facial wrinkle detection

Seong-Gyun Jeong, Yuliya Tarabalka, Josiane Zerubia

MARKED POINT PROCESS MODEL FOR FACIAL WRINKLE DETECTION

Seong-Gyun JEONG, Yuliya TARABALKA, and Josiane ZERUBIA

INRIA, AYIN research team, 2004 route des Lucioles, 06902 Sophia Antipolis, France
E-mails: {firstname.lastname}@inria.fr

ABSTRACT

We propose a new model for wrinkle detection in human faces using a marked point process. In order to detect an arbitrary shape of wrinkles, we represent them as a set of line segments, where each segment is characterized by its length and orientation. We propose a probability density of wrinkle model which exploits local edge profile and geometric properties of wrinkles. To optimize the probability density of wrinkle model, we employ reversible jump Markov chain Monte Carlo sampler with delayed rejection. Experimental results demonstrate that the new algorithm detects facial wrinkles more accurately than a recent state-of-the-art method.

Index Terms— Skin image processing, wrinkle detection, line detection, marked point process, stochastic optimization

1. INTRODUCTION

Since wrinkles are the most significant features of aging, the anti-aging market to treat facial wrinkles is growing rapidly. Cosmetic companies develop various anti-wrinkle creams, and dermatological companies invest in wrinkle filler injections. Thus, there is a need to evaluate performances of such products in that image based wrinkle detection algorithm enables to test large quantity of subjects in an efficient way. Also, vectorized wrinkles can have many applications such as age estimation, forensic application [1] and aged face synthesis.

While the wrinkles are easily distinguishable to human eyes, it is a challenging task for computer vision systems to detect them automatically. First, the wrinkles exhibit distinctive shapes according to ethnic group, gender, age, and personal life style. Moreover, they are observed as subtle features that highly depend on acquisition environment. Batool and Chellappa [2] proposed a stochastic wrinkle detection method based on marked point process (MPP). They employed a second derivative linear filter to extract line structures from an image, and penalized an overlap of wrinkle segments. However, their solution strongly depends on the initial condition, and fails to detect complex patterns of wrinkles.

We propose a wrinkle detection algorithm based on MPP framework. We suppose that wrinkles are interacting line segments that follow the proposed probability density. Specifically, the probability density of wrinkle model is designed to accept smoothly connected lines which are corresponding to edge profile. The reversible jump Markov chain Monte Carlo (RJMCMC) [3] with delayed rejection [4] is employed to maximize this density. Experimental results show that the proposed method provides better wrinkle detection performances when compared to the algorithm of Batool and Chellappa [2].

The authors would like to thank N. Batool for many useful discussions and providing two experimental images. The last author would like to thank Prof. R. Chellappa for the initial contact with his PhD sutdent (N. Batool).

The rest of this paper is organized as follows. Section 2 provides basics of spatial point processes. Section 3 proposes a probability density of wrinkle model. Section 4 describes RJMCMC and acceleration scheme. Section 5 demonstrates experimental results. Finally, conclusions are drawn in Section 6.

2. SPATIAL POINT PROCESSES

This section briefly reviews the spatial point processes [5, 6], on which the proposed model is based. Point processes describe a scene by an unordered set of points in a compact set $\mathcal{F} \subset \mathbb{R}^d$. The set of configurations Ω_n that consists of n points of $\mathcal{F}$ is defined as

$$\Omega_n = \{\mathbf{p}_1, ..., \mathbf{p}_n\}, \mathbf{p}_i \in \mathcal{F}.$$

A point process $\mathbf{X}$ on $\mathcal{F}$ maps a measurable probability space onto the configuration space $\Omega = \cup_{n=0}^{\infty}\Omega_n$.

For example, the homogeneous Poisson process counts the number of points following a discrete Poisson distribution. However, in many applications, points can interact with each other. An MPP introduces additional parameters (mark) for each point to detect objects in a scene. The mark typically encodes the geometric shape of an object. In this work, we assume that a wrinkle corresponds to a line segment, $w_i = (\mathbf{p}_i, \mathbf{m}_i)$, where $\mathbf{p}_i = (x_i, y_i) \in \mathcal{F} \subset \mathbb{R}^2$ is the center of the line in image coordinate space, and $\mathbf{m}_i = (\ell_i, \theta_i)$ is the object parameter vector consists of a length and an orientation, respectively. Furthermore, a probability density $f(\cdot)$ consists of a data fidelity and a prior energy terms. If a point agrees with the probability density function, such a point has a higher chance of being accepted into the final configuration.

3. PROBABILITY DENSITY OF WRINKLE MODEL

This section proposes a probability density of wrinkle model to fit into the given image data, and to constrain spatial relationships of line segments. First, the data term employs the local image gradient. Second, we design the prior energy, which would encourage smooth connection of wrinkle segments while avoiding abrupt intersections of those. The probability density function $f(\cdot)$ can be represented as the product of a data dependent term $h_d(\cdot)$ and a prior energy $h_p(\cdot)$, given by

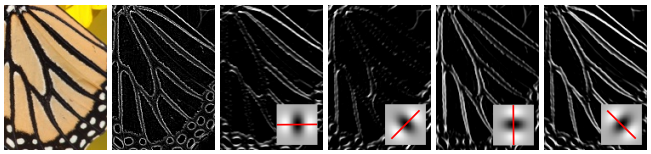
$$f(\mathbf{w}) \propto h_d(\mathbf{w})h_p(\mathbf{w}), \tag{1}$$

where $\mathbf{w} = \{w_i, \cdots, w_n\}$ is a configuration of wrinkle segments.

In this section, we develop the stochastic wrinkle model in a Gibbs energy form $U(\mathbf{w}) \propto -\log f(\mathbf{w})$ which is associated with the probability density function $f(\mathbf{w})$.

3.1. Data likelihood

Wrinkles have a different intensity level compared to normal skin, and they are typically shown as thin lines in an image. A second



(a) Input (b) LoG (c) ∇I_{0° (d) ∇I_{45° (e) ∇I_{90° (f) ∇I_{135°

Fig. 1. When compared to the Laplacian of Gaussian filter responses (b), the steering filter responses (c)–(f) highlight edge components with respect to the specified orientation. Linear filters associated with given orientations are depicted in right bottom of each image.

derivative filter is widely used to extract such linear structures. In addition, we would like to use the orientation information, since we assume wrinkles are oriented line segments. Freeman and Adelson proposed a steerable filter [7] that detects local orientation of the edge. For any arbitrary orientation, a steerable filter can be generated from linear combination of basis filters. In this work, we employ second derivative of Gaussian functions as the basis filter set. The basis filter set for a pixel $\mathbf{p}$ is given by $\mathbf{G}(\mathbf{p}) = \left[\frac{\partial^2 g(\mathbf{p})}{\partial x^2}\ \frac{\partial^2 g(\mathbf{p})}{\partial x \partial y}\ \frac{\partial^2 g(\mathbf{p})}{\partial y^2}\right]^\intercal$, where $g(\mathbf{p})$ denotes the Gaussian function of $\mathbb{R}^2$. Then, the interpolating function of orientation θ is $\mathbf{k}(\theta) = [\cos^2\theta\ -\sin 2\theta\ \sin^2\theta]^\intercal$. Finally, a steerable filter associated with orientation θ is given by

$$g_\theta(\mathbf{p}) = \mathbf{k}^\intercal(\theta)\mathbf{G}(\mathbf{p}). \tag{2}$$

As shown in Fig. 1, steerable filter responses highlight a corresponding orientation. Batool and Chellappa [2] used the Laplacian of Gaussian (LoG) filter to extract thin line structures from the image; however, the LoG filter cannot measure directional information.

We define the data likelihood as the sum of steerable filter responses according to the length and the orientation of a given wrinkle segment w_i:

$$U_d(w_i) = -\sum_{\mathbf{p}\in w_i} \nabla I_{\theta_i}(\mathbf{p}), \tag{3}$$

where θ_i is an angle corresponding to the wrinkle segment w_i, I denotes the grayscale input image, and $\nabla I_{\theta_i} = g_{\theta_i} * I$ is the gradient intensity obtained by the steerable filter g_{θ_i}.

3.2. Prior energy

The prior energy of the proposed model defines the relationship of adjacent line segments. Since the wrinkles appear on skin as curved lines with a small curvature, the prior model encourages smoothly connected line segments to be accepted into the final configuration. On the other hand, a repulsive interaction is defined to reject abrupt intersection and overlapping of segments.

Let $\mathcal{N}(w_i)$ be a neighborhood set of lines for a wrinkle w_i. We assume two lines to be neighbors when the distance between their centers is smaller than the sum of their half lengths:

$$\mathcal{N}(w_i) = \left\{ w_j \in \mathbf{w} \;\middle|\; 0 < \|\mathbf{p}_i - \mathbf{p}_j\|_2 \le \frac{\ell_i + \ell_j}{2} \right\}. \tag{4}$$

The $\mathcal{N}(w_i)$ can be partitioned into two disjoint sets: attractive neighborhood set $\mathcal{N}_a(w_i)$ and repulsive neighborhood set $\mathcal{N}_r(w_i)$. For simplicity, we only describe how to build the attractive neighborhood set because the repulsive neighborhood set is trivial, $\mathcal{N}_r = \mathcal{N} \setminus \overline{\mathcal{N}_a}$.

The attractive neighborhood of w_i is composed of connected and aligned line segments at the ends of the w_i. That is

$$\mathcal{N}_a(w_i) = \{ w_j \in \mathcal{N}(w_i) \mid d_{ij} \le \tau_\epsilon, \theta_{ij} \le \tau_\theta \}. \tag{5}$$

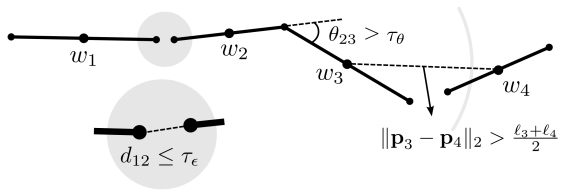


Fig. 2. Illustration of a wrinkle configuration that includes an attractive interaction $\mathcal{N}_a(w_1) = \{w_2\}$, a repulsive interaction $\mathcal{N}_r(w_2) = \{w_3\}$, and a single segment $\mathcal{N}(w_4) = \emptyset$.

If two line segments share their end points at the same image coordinate, then they are said to be connected. Also, we assume that two lines are connected if the distance of the end points is within the predefined distance τ_ϵ. The distance of the end points d_{ij} is computed as

$$d_{ij} = \min\left\{ \|\mathbf{p}_i^L - \mathbf{p}_j^R\|_2,\ \|\mathbf{p}_i^R - \mathbf{p}_j^L\|_2 \right\}, \tag{6}$$

where $\mathbf{p}^L$ and $\mathbf{p}^R$ denote the coordinates of the left and the right end points, respectively. Moreover, a curvature of connected lines should be minimized to avoid jitter of wrinkles. The curvature is simply measured by the angular difference of the connected lines:

$$\theta_{ij} = \min\{|\theta_i - \theta_j|, \pi - |\theta_i - \theta_j|\}. \tag{7}$$

Intuitively, a smaller value of θ_{ij} represents a smaller curvature. Note that w_j will be excluded from $\mathcal{N}_a(w_i)$ when the angular difference θ_{ij} is greater than the pre-defined angle τ_θ. After constructing $\mathcal{N}_a(w_i)$, we can compute the attractive energy $U_a(w_i)$ given by

$$U_a(w_i) = \sum_{w_j \in \mathcal{N}_a(w_i)} \left(1 - \frac{\phi(d_{ij}, \tau_\epsilon) + \phi(\theta_{ij}, \tau_\theta)}{2} \right), \tag{8}$$

where $\phi(x, \tau) = \frac{1}{\tau^2}\left(\frac{\tau^2 - x^2}{1 + x^2}\right)$ is the *Quality Function* [8] which maps a bounded input value $x \in [-\tau, \tau]$ onto $[0, 1]$.

We penalize a wrinkle segment according to the cardinality of the repulsive neighborhood set. Then, the repulsive prior energy $U_r(w_i)$ is:

$$U_r(w_i) = n(\mathcal{N}_r(w_i)). \tag{9}$$

Also, a single segment which has no neighborhood is penalized with a positive constant value. This is because such a single segment corrupts the final configuration with fragments.

For better understanding, we provide graphical interpretation of the neighborhood system in Fig. 2. We can observe an attractive interaction between w_1 and w_2. The w_3 is included in the repulsive neighborhood set of w_2, because the angular difference θ_{23} exceeds the threshold value. The w_4 is apart from the other line segments; thus, w_4 is a single segment.

To sum up, the prior energy $U_p(w_i)$ is given by

$$U_p(w_i) = \begin{cases} \beta_1 & \text{if } \mathcal{N}(w_i) = \emptyset, \\ \beta_2 U_a(w_i) + \beta_3 U_r(w_i) & \text{if } \mathcal{N}(w_i) \neq \emptyset, \end{cases} \tag{10}$$

where $\beta_{k,k=\{1,2,3\}}$ represent weighting factors which control the relative importance of each term. β_1 is a threshold to control the birth of a new wrinkle segment which depends on the resolution and brightness of an image. If β_1 is fixed to a high value, then the final configuration suffers from under-detection. In this work, we set β_1 as a mean of the LoG filter responses multiplied by the minimum length of wrinkle segment $\ell_{\min}$. For attractive interaction, we use $\beta_2 = 0.2\beta_1$. β_3 is set to a large number to realize hard-core condition [5].

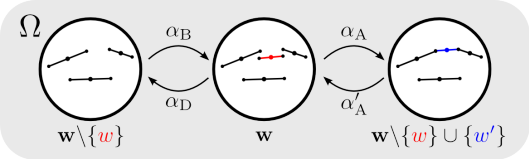


Fig. 3. A state of the Markov chain is related to a wrinkle configuration, and updates its state according to the acceptance ratio of transition kernel.

4. OPTIMIZATION

In this section, we present a stochastic optimization based on simulated annealing and reversible jump Markov chain Monte Carlo (RJMCMC) with delayed rejection. Our goal is to maximize the proposed density (equivalent to minimizing the associated Gibbs energy) in order to extract wrinkles from an image. Since the density function is nonconvex and lies in varying dimension, it is impossible to maximize the given function directly. To solve the optimization problem, we employ simulated annealing method, which optimizes an unnormalized density function $f(\mathbf{w})^{\frac{1}{T}}$ instead of $f(\mathbf{w})$. T is the temperature parameter which gradually decreases from the initial temperature T_0 to zero as the iteration procedure goes on.

4.1. RJMCMC sampler

The RJMCMC algorithm [3] simulates a discrete Markov chain $(X_t)_{t\in\mathbb{N}}$ over the configuration space Ω. The Markov chain converges toward the stationary target distribution which is specified by the probability density of wrinkle model (1). At each iteration, a current configuration $\mathbf{w}$ is locally perturbed by adding, removing, or modifying a wrinkle segment (see Fig. 3). More specifically, a new configuration $\mathbf{w}'$ is proposed according to a transition kernel $\mathcal{Q}(\mathbf{w},\cdot)$ which can be decomposed into different types of moves $\mathcal{Q}_{\mathrm{m}}$ with probability p_{m}. That is

$$\mathcal{Q}(\mathbf{w},\cdot)=\sum_{\mathrm{m}} p_{\mathrm{m}}\mathcal{Q}_{\mathrm{m}}(\mathbf{w},\cdot). \tag{11}$$

For the reversibility of the Markov chain, proposed sub-transition kernels must satisfy the detailed balance condition [3]. The acceptance ratio for the m-th type of move is defined by

$$\alpha_{\mathrm{m}}(\mathbf{w},\mathbf{w}')=\min\left(1,\ \frac{\mathcal{Q}_{\mathrm{m}}(\mathbf{w}',\mathbf{w})}{\mathcal{Q}_{\mathrm{m}}(\mathbf{w},\mathbf{w}')}\frac{f(\mathbf{w}')}{f(\mathbf{w})}\right). \tag{12}$$

Birth-and-Death

The birth kernel adds a new wrinkle segment into the configuration. A new wrinkle segment w is randomly chosen from the considered mark space. The acceptance ratio of birth kernel is given by

$$\alpha_{\mathrm{B}}=\min\left(1,\ \frac{p_{\mathrm{D}}}{p_{\mathrm{B}}}\frac{\lambda(\mathcal{F})}{n(\mathbf{w})+1}\frac{f\left(\mathbf{w}\cup\{w\}\right)^{\frac{1}{T}}}{f(\mathbf{w})^{\frac{1}{T}}}\right), \tag{13}$$

where p_{B} and p_{D} denote the probability of choosing birth and death, respectively. $n(\mathbf{w})$ is the number of segments in the configuration $\mathbf{w}$, and $\lambda(\mathcal{F})$ is the Lebesgue measure over the mark space. That is $\lambda(\mathcal{F})=\mathcal{F}\times|\ell_{\max}-\ell_{\min}|\times|\theta_{\max}-\theta_{\min}|$, where $\ell_{\max}$ is the maximum length of a wrinkle segment. $\theta_{\max}$ and $\theta_{\min}$ denote the maximum and minimum angle of a wrinkle segment, respectively.

In opposite case, the death kernel removes a wrinkle segment from the configuration if the segment disagrees with the proposed density. The acceptance probability of the death kernel is obtained by the inverse of the birth kernel:

$$\alpha_{\mathrm{D}}=\min\left(1,\ \frac{p_{\mathrm{B}}}{p_{\mathrm{D}}}\frac{n(\mathbf{w})}{\lambda(\mathcal{F})}\frac{f\left(\mathbf{w}\backslash\{w\}\right)^{\frac{1}{T}}}{f(\mathbf{w})^{\frac{1}{T}}}\right). \tag{14}$$

Affine transform

Unlike the birth-and-death kernel, the affine transform kernel does not change the cardinality of the current configuration. The affine kernel first selects an arbitrary segment from the current configuration, then updates the object parameters including center position, length and orientation of a line. Therefore, a wrinkle segment w becomes w' by adding an affine transform vector $\mathbf{A}=[\Delta\mathbf{p}\ \Delta\mathbf{m}]^{\intercal}$, where $\Delta\mathbf{p}=(dx,dy)$ and $\Delta\mathbf{m}=(d\ell,d\theta)$, which is an infinitesimal change within the sampling space. That is

$$w'=w+\mathbf{A}. \tag{15}$$

The acceptance ratio of the affine transform kernel is defined by

$$\alpha_{\mathrm{A}}=\min\left(1,\ \frac{f\left(\mathbf{w}\backslash\{w\}\cup\{w'\}\right)^{\frac{1}{T}}}{f(\mathbf{w})^{\frac{1}{T}}}\right). \tag{16}$$

A new configuration is accepted as a new state of the Markov chain according to the acceptance ratio. A higher probability will be drawn if a candidate line segment fits the proposed density.

4.2. RJMCMC with delayed rejection

Although the affine transform kernel reduces optimization time, random-walk method often returns to the previous configuration and hence delays the convergence. We also observe that the visualized result shows punched lines due to the conflict of continuous sampling space and discrete image coordinates. To overcome these issues, we perform the delayed rejection scheme [4].

Once a new wrinkle segment w' proposed by the affine transform kernel is rejected, we give it a second chance to be accepted into the configuration by enforcing the connectivity. For each end of w', we search for the nearest end point among the segments of the current configuration within the predefined vicinity. In this work, we set the search range within a 6-pixel-radius circle centered at each of the endpoints of w'. If such nearest ends exist, we propose an alternative line segment w^* by interpolating the retrieved points. Obviously, w^* is firmly connected with the adjacent lines, and its object parameters are resampled according to the end points. The acceptance ratio of delayed rejection is defined by

$$\alpha_{\mathrm{A}}^2(\mathbf{w},\mathbf{w}',\mathbf{w}^*)= \min\left(1,\ \frac{\mathcal{Q}_{\mathrm{A}}(\mathbf{w}^*,\mathbf{w}')}{\mathcal{Q}_{\mathrm{A}}(\mathbf{w},\mathbf{w}')}\frac{\mathcal{Q}_{\mathrm{A}}^2(\mathbf{w}',\mathbf{w}^*,\mathbf{w})}{\mathcal{Q}_{\mathrm{A}}^2(\mathbf{w},\mathbf{w}',\mathbf{w}^*)}\frac{[1-\alpha_{\mathrm{A}}(\mathbf{w}^*,\mathbf{w}')]}{[1-\alpha_{\mathrm{A}}(\mathbf{w},\mathbf{w}')]}\frac{f(\mathbf{w}^*)}{f(\mathbf{w})}\right), \tag{17}$$

where $\mathbf{w}'=\mathbf{w}\backslash\{w\}\cup\{w'\}$ and $\mathbf{w}^*=\mathbf{w}\backslash\{w\}\cup\{w^*\}$. $\mathcal{Q}_A^2$ is the second phase transition kernel for delayed rejection [4]. Due to the symmetric property of the transition kernel, (17) is simplified as

$$\alpha_{\mathrm{A}}^2=\min\left(1,\ \frac{f(\mathbf{w}^*)^{\frac{1}{T}}-f(\mathbf{w}')^{\frac{1}{T}}}{f(\mathbf{w})^{\frac{1}{T}}-f(\mathbf{w}')^{\frac{1}{T}}}\right). \tag{18}$$

Table 1. Quantitative evaluation of [2] and the proposed algorithm.

	Batool and Chellappa			Proposed		
	r_{detect}	r_{false}	F	r_{detect}	r_{false}	F
B3	0.72	0.10	0.29	0.71	0.03	0.31
BC4	0.78	0.02	0.27	0.83	0.02	0.28
SA	0.57	0.05	0.18	0.77	0.02	0.27
MF	0.63	0.02	0.27	1.00	0.02	0.31

5. EXPERIMENTAL RESULTS

We demonstrate the performances of the proposed algorithm on four test sequences with different spatial resolutions. Two of them are collected on the Internet, and two other sequences are the courtesy of Batool and Chellappa [2]. For all test sequences, we choose the parameters as follows: $\ell_{\min} = 3$ pixels and $\ell_{\max} = 14$ pixels. The orientation θ is varying from -90° to 90° by 1° of increments. The marginal distance of connected segments τ_ϵ is fixed to 1.2 pixels, and the maximum angular difference of aligned segments τ_θ is 30°.

The initial configuration is set up according to the highest LoG filter responses similar to [2]. We perform the simulated annealing algorithm to perturb the Markov chain. The temperature T is initially set to 120, and it decreases according to the cooling schedule $T_t = \frac{T_0}{\log(1+t)}$ as the iterations go on. The algorithm is terminated if the number of iterations reaches 5.0×10^5.

Fig. 4 compares the detection results of the proposed method with those obtained by applying Batool and Chellappa's technique [2]. Since the algorithm [2] prohibits to remove seed segments, its output contains wrinkles with unnatural orientation. While the algorithm of Batool and Chellappa shows jitters in the final configuration, the proposed algorithm yields smooth wrinkles for all test sequences.

For the quantitative evaluation, we employ the same measures used in [2]: r_{detect} shows correctness of wrinkle detection, and r_{false} computes the over-detection ratio. We also provide the F-measure [9] which is independent of image size. The F-measure is defined by

$$F = \frac{2PR}{P+R},$$

where P and R denote the precision and recall, respectively. Table 1 summarizes the quantitative performance of the proposed algorithm and of [2], respectively .

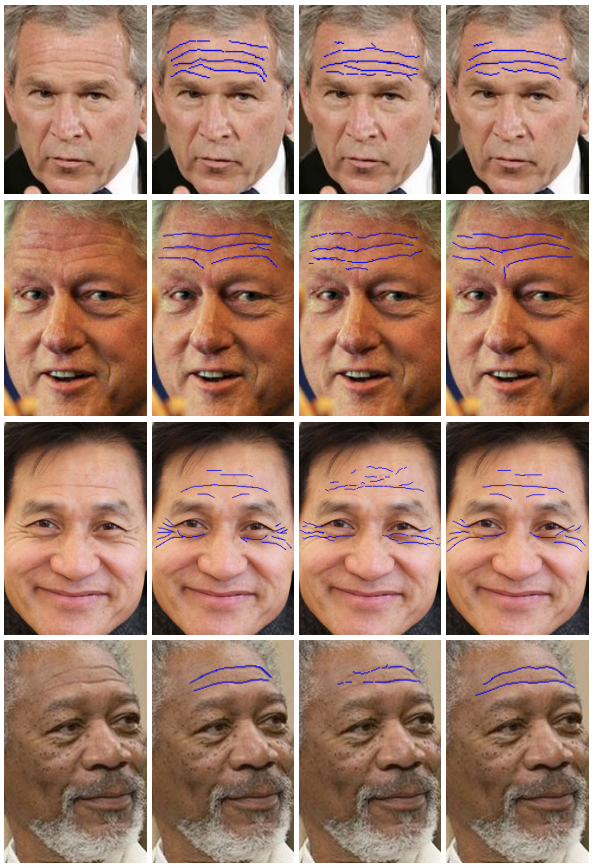
6. CONCLUSIONS

We developed a new marked point process model for facial wrinkle detection. We represented wrinkles as vectorized lines, and proposed a probability density using image gradient and prior knowledge of wrinkle shapes. We developed transition kernels for RJMCMC sampler and delayed rejection scheme. Simulation results confirm that the proposed MPP model is less dependent on the initial condition and shows higher performances when compared to the recent state-of-the-art model of Batool and Chellappa [2].

7. REFERENCES

[1] N. Batool, S. Taheri, and R. Chellappa, "Assessment of facial wrinkles as a soft biometrics," in *Proc. IEEE FG*, Apr. 2013, pp. 1–7.

[2] N. Batool and R. Chellappa, "Modeling and detection of wrinkles in aging human faces using marked point processes," in *Proc. ECCV Workshops*, Oct. 2012, pp. 178–188.



(a) Input (b) Manually labeled (c) Batool and Chellappa (d) Proposed

Fig. 4. Comparison of the proposed algorithm with [2]. The name of sequences is top to bottom: B3, BC4, SA, and MF. The manually labeled images of B3 and BC4 are courtesy of N. Batool.

[3] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 771–732, 1995.

[4] P. J. Green and A. Mira, "Delayed rejection in reversible jump Metropolis-Hastings," *Biometrika*, vol. 88, no. 4, pp. 1035–1053, 2001.

[5] M. N. M. van Lieshout, *Markov point processes and their application*, Imperial College Press, London, 1st edition, 2000.

[6] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*, John Wiley & Sons, Inc., 1987.

[7] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE TPAMI*, vol. 13, no. 9, pp. 891–906, Sept. 1991.

[8] C. Lacoste, X. Descombes, and J. Zerubia, "Point processes for unsupervised line network extraction in remote sensing," *IEEE TPAMI*, vol. 27, no. 10, pp. 1568–1579, Oct. 2005.

[9] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE TPAMI*, vol. 26, no. 5, pp. 530–549, May 2004.